

Locality oriented feature extraction for small training datasets using non-negative matrix factorization

Khoa Dang Dang · Thai Hoang Le

Received: 30 November 2013 / Accepted: 16 July 2014 / Published online: 6 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract This paper proposes a simple and effective method to construct descriptive features for partially occluded face image recognition. This method is aimed for any small dataset which contains only one or two training images per subject, namely Locality oriented feature extraction for small training datasets (LOFESS). In this method, gallery images are first partitioned into sub-regions excluding obstructed parts to generate a collection of initial basis vectors. Then these vectors are trained with Non-negative matrix factorization algorithm to find part-based bases. These bases finally build up a local occlusion-free feature space. The main contribution in this paper is the incorporation of locality information into LOFESS bases to preserve spatial facial structure. The presented method is applied to recognize disguised faces wearing sunglasses or scarf in a control environment without any alignment required. Experimental results on the Aleix-Robert database show the effectiveness of the LOFESS method.

Keywords Disguised face recognition · Partial occluded face recognition · Non-negative matrix factorization · Alignment free face recognition

1 Introduction

Human face recognition has been long studied in the research community with many achievements [1, 2]. It plays an important role in security, supervision, human–machine interaction

and more. Face images offer an advantage over other biometric features that it is far more easy to be captured with the help of digital cameras increasingly popular nowadays. For human, it is not so difficult to recognize people in many conditions. But for computers, there are many challenges still troubling researchers.

One problem that draws much of attention is recognizing a partially occluded face. The occlusion is caused by a facial accessory such as sunglasses or scarf [3]. This is also called disguised face recognition. A common solution is to focus on the feature representation so that discriminative information is effectively extracted. In addition, it is not always possible to acquire many photos of each person easily. In practice, some applications requiring this feature space is efficiently built based on a small training dataset, which means only one or two subject's images are available. This is also one of the main concerns in this paper.

The disguised face recognition has different approaches. Many of the state of the art methods, such as SRC [4] and RSC [5], utilize the redundant information based on the availability of large scale image galleries. This condition is unfeasible in some applications when only a very few number (one or two) of training images are available. In another approach, non-negative matrix factorization (NMF) based methods [9, 10] show promising results when applying to small training datasets [14] due to their ability to learn part-based features naturally. However, these methods just focus to control the sparseness of NMF features, while spatial relationship information among bases is not exploited sufficiently. This paper concentrates on the problem of building an occlusion-excluded feature space for recognizing partial occluded faces, such as by wearing eyeglasses or scarves, based on a small gallery set, namely Locality oriented feature extraction for small training datasets (LOFESS). Each subject in the dataset has one or two images

K. D. Dang · T. H. Le (✉)
Department of Information Technology, University of Science,
227 Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam
e-mail: lhthai@fit.hcmus.edu.vn

K. D. Dang
e-mail: ddkhoa@fit.hcmus.edu.vn

captured in a controlled environment (straight faces with neutral expression and balanced light condition), without any alignment needed. Moreover, spatial information is explicitly employed to enhance the robustness to occlusion. Noted that this method can be extended for other types of disguises.

LOFESS first requires the disguise condition to be identified manually or automatically. It is assumed the occlusion detection step, which is out of the scope of this paper, has been done by another algorithm or by a user. Then, gallery images are split into suitable regions to construct an initial basis set. These bases are designed so that none of any pixel in the detected occluded area is involved. It is important and reasonable to remove these pixels because they certainly degrade the recognition performance. The next step is training these bases into localized facial components by Non-negative matrix factorization. Basically, these components are matrices with all the entries are greater or equal to zero. This enable them to mutually combine together to reconstruct original faces. As a contribution, a splitting strategy is designed to incorporate spatial relationship into these components. Finally, occlusion-free bases are matched to identify the target.

Figure 1 summarizes the mentioned steps in this paper. To show the effectiveness of the proposed LOFESS method, we use a subset of the Aleix-Robert database [11] which is standard in many related research. This dataset offers a large amount of face images of 100 people wearing sunglasses or scarves which is a standard for experiments and comparison. The remainder of this paper is organized as follows. In Sect. 2, we highlight the main studies in this problem. Section 3 describes in detail our feature space construction LOFESS method following by the comparison with state of the art algorithms. Experimental methodology and results are presented in Sect. 4. Finally, we make a conclusion and propose future works in Sect. 5.

2 Backgrounds

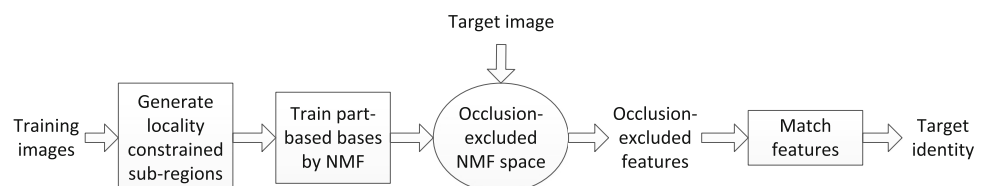
This section mainly reviews the recent literature of feature representation for disguised face recognition. Features could be extracted at various scale from a whole face to small pixel blocks over the image and represented by code-based or subspace-based methods.

Intuitively, partial face occlusion significantly degrades the recognition performance. A possible approach is to recover these parts before recognizing who they are. Chiang and Chen's solution [6] automatically detects occlusion and recovers the occluded parts. At the end, the whole face is matched with faces recovered from person-specific PCA [12] eigenspaces after a gradual illumination adjustment process. As authors' discussion, this model depends heavily on manually fitting active appearance model (AAM) [13] landmarks on each input faces which is not reliable when eye region is covered.

Instead of recovering, most of recent arts choose to remove occluded parts and extract local features from the rest of the image. Code-based approaches have been widely investigated in the literature due to their high recognition performance. The main idea is to approximate original data through linear combination of only a few (sparse) coding basis, or atoms, chosen from an over complete dictionary. Wright et al. [4] recently proposed the sparse representation based classification (SRC) scheme for face recognition which achieved impressive performance. Images are split into a grid of smaller regions and applying SRC separately. Each block is treated as an atom without any projection into a subspace or feature extraction. Their method shows high robustness to face occlusion. Starting from this success, many variants of SRC make further improvements. Nguyen et al. [7] built a multi-scale dictionary. In their work, each image is scaled by 2 four times and split into 16, 8, 4 and 2 blocks, respectively, at each level. SRC is then performed on separated group of blocks. Yang and Zhang [15] integrated an additional occlusion dictionary. The built-in atoms are extracted from image local Gabor features [16] to enhance the compactness and reduce the computational cost of sparse coding. A separated block will be removed if it is classified as occluded or taken into account if it is non-occluded. These methods use simple voting strategies to fuse the recognition result from separated blocks so the spatial relationship among these blocks are not considered properly.

In the approach of combining sparse coding with global representation, Yang et al. [5] based on the maximum likelihood estimation principle to code an input signal by sparse regression coefficients. This method utilizes an iterative process to create a map weighting occluded and non-occluded pixels differently. The weighted input image is then

Fig. 1 An overview of face recognition based on our LOFESS method



matched with template images in the dictionary. Zhou et al. [17] included the Markov Random Field model to identify and exclude corrupted regions from the sparse representation. This method can even iteratively reconstructed an input face from un-occluded part. Liao and Jain [18] proposed an alignment-free approach based on a large scale dictionary of SIFT descriptors. The disadvantage of all sparse-based methods in this problem context is a large number of gallery images must be obtained in advance to build dictionaries.

Non-negative matrix factorization (NMF) [9,10,19] is another approach which has been proven a useful tool for decomposing data into part-based components. These components are non-negative meaning all elements in factorized matrices are greater than or equal to zero. This idea comes from biological modeling research aiming to simulate receptive fields of human visual system where input signals are mutually added (not canceled each one out). One important property of NMF is it naturally results in sparse features which are highlighted salient local structures from the input data. This property is valuable when dealing with occlusion and dimension reduction. Showing that sparseness of NMF bases is somehow a side effect, Hoyer [20] introduced a constraint term to explicitly control the degree of sparseness of learned bases.

With the same purpose, Hoyer and Shastri [21,22] imposed non-negativity as a constraint in sparse coding model called Non-negative sparse coding (NNSC). This method pursuit sparseness and part-based representation at the same time. However, as we observed, the constraint is not enough to guarantee both properties simultaneously. Hoyer [20] has the same conclusion about the trade-off between sparsity, localization and data representation sufficiency. In these methods, learned bases converge randomly because there is no constraint on each facial part position. This shortcoming results in a waste of features and ineffectiveness when recognizing disguised faces not only because these features have nothing to deal with occluded regions but also degrade the recognition performance. To tackle this problem, Oh et al. [14] divided input images into non-overlapped patches to detect occlusion. Then, the matching is performed in the Local non-negative matrix factorization (LNMF) space [23] constructed by the selected occlusion-free bases.

Apart from the above discussed methods, there are various approaches base on face sub-images such as Martinez's probabilistic approach [25] which is able to compensate for partially occlusion, Ekenel and Stiefelhagen's alignment-based approach [24] resulted from Rentzeperis et al. [26] that registration errors have dominant impact on recognition performance over the lacking of discriminative information.

3 LOFESS: an effective and efficient feature representation for small training datasets

3.1 Face sub-regions with spatial relationship preserving constraints

This section proposes a new face sub-region representation to construct inputs for training by NMF in the next step, which are incorporated with spatial constraints at the same time.

This paper mainly deals with faces wearing sunglasses or scarves, but note that the same strategy could apply for other type of partial disguise.

The main point is to build a feature representation without taking any pixels in the eyes and mouth regions. However, these two regions are thought to carry most of identifying features of human face. Our aim is to exclude them but not affect or even boost the recognition performance. This could be achieved by employing spatial relationship to complement for the loss of information.

Input:

- A dataset consisting of m images at the same size of $p \times q$
- n is the number of basis vectors we wish to receive after training
- Information about occlusion (i.e. which part need removing) R

Loop for k from 1 to n

Choose one image I from the dataset randomly

Construct a new image

$$I'_{ij} = \begin{cases} I_{ij}, & r_1 \leq i \leq r_2, 1 \leq j \leq q \\ 0 & \end{cases}$$

Choose $1 \leq r_1 < r_2 \leq p$ so that the image I' will not contain any pixel in the occluded regions R .

Transform I' into the column vector $w_k \in \mathbb{R}^{d \times 1}$, with $d = p \times q$

End loop

Output:

- The matrix $W_0 \in \mathbb{R}^{d \times n}$ including column vectors w_i .

In this data preparation step, information about occlusion could be supplied by a user or a result from an occlusion detection algorithm. This step acts as a guidance for features to converge into regions outside the occlusion, eyes or mouth, and just focus to extract information from other parts. Figures 2 and 3 show some sample bases before and after training. The top row (a) depicts original images I . The second row (b) is initial basis images I' with regions split from (a). The bottom row (c) are bases learned from (b), i.e. W^* (will be

Fig. 2 For recognizing subjects wearing sunglasses: from original images (a), initial basis images (b) are constructed, and final LOFESS bases (c) are learned with eye regions removed

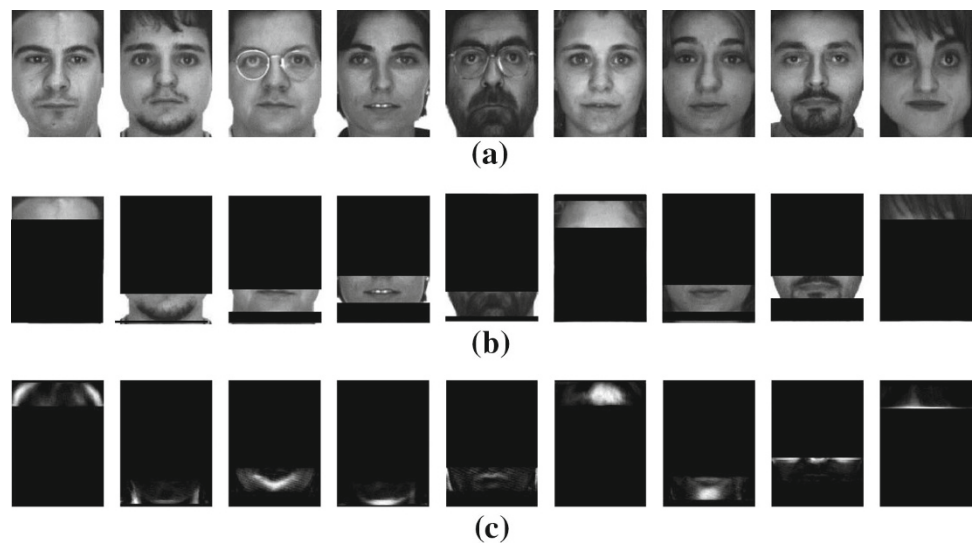
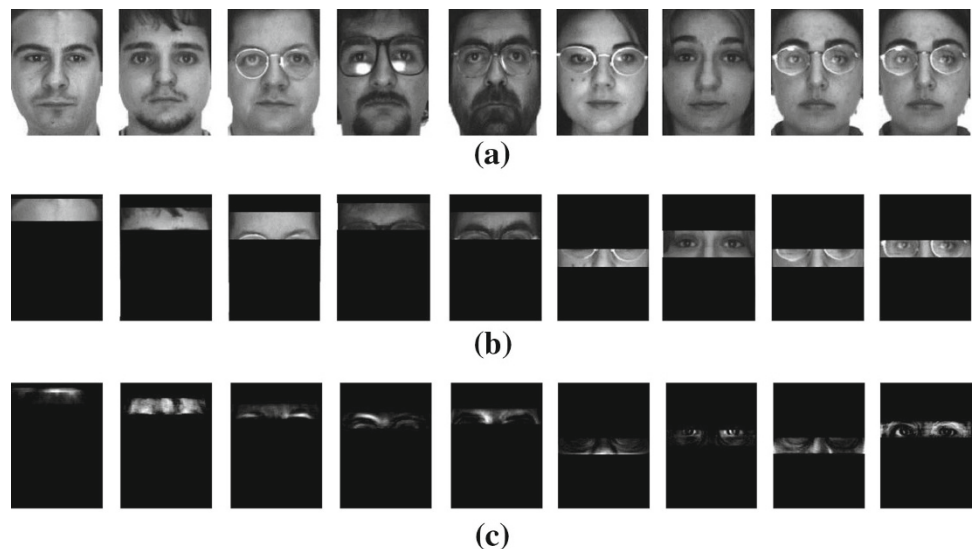


Fig. 3 For recognizing subjects wearing scarves: from original images (a), initial basis images (b) are constructed with mouth regions removed to learn final LOFESS bases (c)



presented in the next section). These regions depend on the choice of r_1 and r_2 so that their combination could cover an entire face excluding occluded areas. Note that these figures were chosen randomly for illustration purpose, there is no correspondence between them.

The use of occluded regions could result in performance degradation. State of the art methods have employed different approaches to remove or avoid occlusion. LOFESS improves this idea by both zeroing out any pixel in occluded areas and preserving the facial structure at the same time. It means each LOFESS basis carries robust, complementary information (which person and which corresponding facial part) for recognition. Also note that only this step requires occlusion forms to be identified in advance. When matching, a testing image will be represented just based on available trained bases, none of which corresponds to occluded areas.

So occlusion is removed naturally without any additional computation.

If there is no or minor occlusion that could be neglected, all facial regions are taken in to account. The problem then becomes recognizing faces without occlusion and the algorithm is still applied properly.

3.2 Training occlusion-free part-based features with NMF

The NMF aims to learn part-based representation of faces. Let V is a column vector matrix, each column represents an image in the training dataset. This method tries to find basis vectors W and coefficients H that best approximate V , i.e. minimize the error:

$$\varepsilon = \|V - WH\| \quad (1)$$

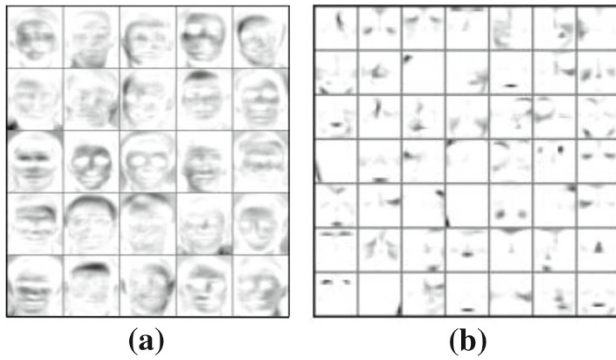


Fig. 4 Example basis vectors learned by the original NMF

with the constraint of non-negative on W and H (all negative values will be assigned by zeros during computation). The optimal solution for W and H is given by iterating the following Multiplicative Update Rule algorithm [9]. The iteration stops when ε lower than a predefined threshold or after a certain number of update times.

$$H_{au} = H_{au} \frac{(W^T V)_{au}}{(W^T W H)_{au}} \quad (2)$$

$$W_{ia} = W_{ia} \frac{(V H^T)_{ia}}{(W^T H H^T)_{ia}} \quad (3)$$

with a, u, i are row and column indexes.

Originally, H and W are randomly generated. However, in practice, it doesn't guarantee bases will converge to local parts as expected and usually results in global representation [20, 27] (Fig. 4a).

LOFESS initializes W from W_0 in Sect. 3.1. This method differs from Hoyer's [20], called Non-negative sparse coding (NNSC), which tried to control the localization of W and sparseness of H at the same time. NNSC is not able to decide which local part on a face to focus on. In Fig. 4b from Hoyer's paper, these features converged randomly to any part. For instance, a region around one's eyes is useless when recognizing a person wearing sunglasses.

3.3 Face recognition with locality constrained features

We improved the model from Shastri and Levine [22] by adding spatial constraint in the feature extraction phase (Fig. 5).

3.3.1 Training

From the initial dataset $D \in \mathbb{R}^{p \times q \times m}$ consisting of m images at the same size $p \times q$, we construct the matrix $V \in \mathbb{R}^{d \times m}$, $W_0 \in \mathbb{R}^{d \times n}$ and initialize a matrix $H_0 \in \mathbb{R}^{n \times m}$ with random values, $d = p \times q$.

NMF takes V , W_0 and H_0 as the inputs. After the training, we will receive the optimal bases W^* and coefficients H^* . Together, $W^* H^*$ best approximates the training set V . Figures 2c and 3c depict some samples of W^* , note that none of them relates to occluded areas.

The feature space W^+ is constructed and each column vector v_k in V is projected on this space to obtain a feature vector h_i

$$W^+ = (W^{*\top} W^*)^{-1} W^{*\top} \quad (4)$$

$$h_i = W^+ v_i, \quad i = 1 \dots m \quad (5)$$

In some practical situations, a person may wear sunglasses, a scarf, both or anything else. Depend on occlusion types, several corresponding W^+ could be constructed in advanced.

3.3.2 Matching

Let $y \in \mathbb{R}^{d \times 1}$ represents an image of an unidentified subject wearing disguise. Based on the form of disguise, identified by a user or an algorithm, e.g. sunglasses or a scarf, the corresponding W^+ is chosen. Project y onto the feature space W^+ to receive vector h_y

$$h_y = W^+ y \quad (6)$$

The subject is assigned to the nearest neighbor class based on the Euclidean distance from h_y to all h_i of training images. It means find

$$k = \min_i d(h_y, h_i) = \min_i L_2(h_y, h_i), \quad \text{with } i = 1 \dots m \quad (7)$$

In conclusion, y belongs to the same class of v_k .

The matching process is illustrated in Figs. 6 and 7. Training (a) and testing (e) images are projected on W^+ (b and f are the same) to produce feature vectors h_i and h_y (c and g). The feature vectors impose representation (d and h) of input images based on non-occluded bases.

3.4 Merits of LOFESS

The proposed method LOFESS has the following merits in the small training dataset context. Firstly, LOFESS is robust to various types of partial occlusion. It transforms a disguised face into the occlusion-excluded LOFESS space and perform the matching only on visible parts. The strength of LOFESS is spatial relationship is preserved to complement for losing information in occluded parts. Secondly, LOFESS achieves high recognition performance on small training datasets because it exploits both global and local information from limited resources. Each basis corresponds

Fig. 5 Training and matching process

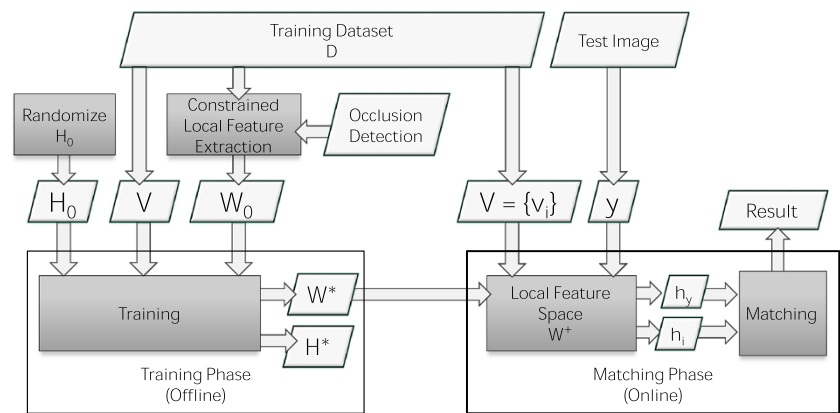


Fig. 6 Matching between training and testing samples in the sunglasses dataset

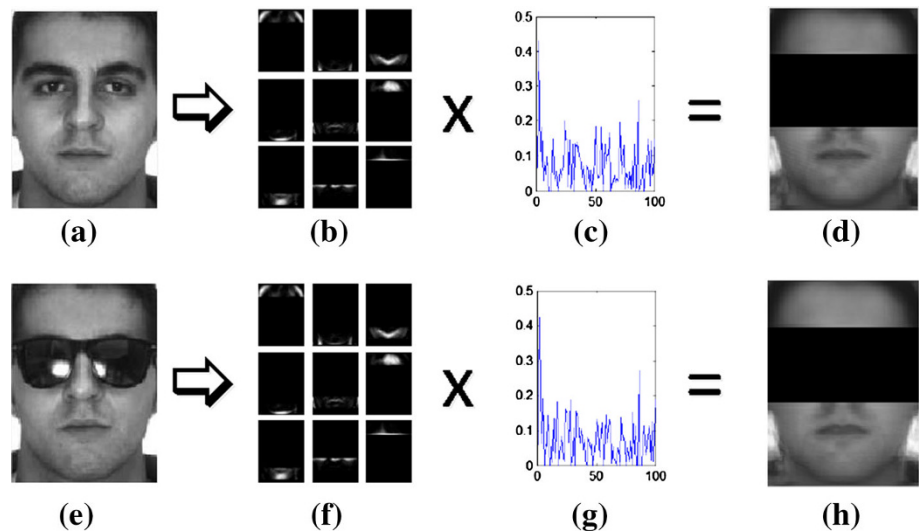
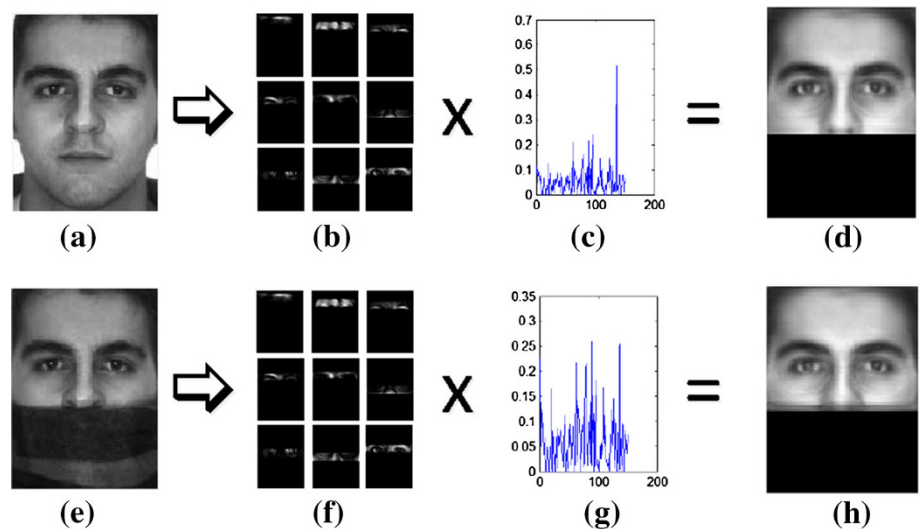


Fig. 7 Matching between training and testing samples in the scarf dataset



to a facial part and its relative position to the whole face structure implies spatial relationship. Indeed, within a single basis, the meaningful information (nonzero pixels) concentrates in a small region. In this paper, we keep the whole image for easy visualization and interpretation. When implementing,

a suitable data structure could be employed to reduce the number of dimensions by dismissing or compressing blank (black) regions. Thirdly, LOFESS is easily incorporated with prior knowledge from occlusion detection algorithms or from a user in semi-supervised applications. Automatic detection

Table 1 Comparison between LOFESS and other methods

	Sparseness	Locality	Minimum number of training images
SRC	On coefficients	Block partitioning sparse error term	8 images/person
RSC	On coefficients	Sparse error term	4 images/person
NMF	On bases	Spatially localized	1 images/person
SLNMF	On bases	Spatially localized	1 images/person
LOFESS	On bases	Spatially localized + structure constraint preserving images/person	1 images/person

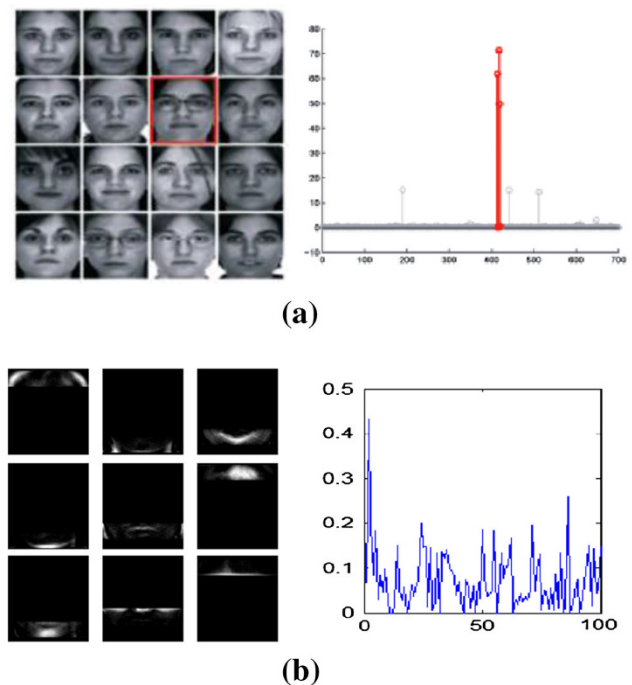
are not readily applied in practice and cost more computation. Meanwhile, supervising applications are usually monitored by users. LOFESS only requires a user to mark occluded region in a template image in the beginning. The template is then applied to all images and no need any user interaction afterward. This way of manipulation is easy and fast for users as well as support the system reliability.

3.5 Comparison with existing methods

LOFESS can be considered as a method for learning sparse features with locality constraints to construct an occlusion-free feature space. At first, constraints are applied on original data regarding to occlusion types. After that, this data becomes input to an iterative training process to learn part-based bases. These bases form a subspace on which an input face is projected to find a occlusion-excluded representation suitable for small training datasets. In this section, LOFESS is compared with two representative approaches based on the same sparseness property as summarized in Table 1.

SRC and variants (e.g. RSC) seek for sparse combination of bases, which means choosing a set of coefficients with very few elements greater than zero. In return, bases are dense to produce enough information for recognition. To achieve robustness to occlusion, these bases are split into a grid or selective regions. Each region is treated separately and results are fused by voting. This doesn't take into account the spatial relationship between regions. Additional sparse error term is integrated to overcome this drawback but consumes more time and computation. As reported in authors' paper [4], it took more than 20 s to process one image. Moreover, the number of gallery images needed to reach the optimal performance is more than the assumption in this problem, which is one or two training images per person.

NMF-based methods, on the other hand, try to learn sparse bases and combination of these bases to represent input faces. The spatially localized bases enhance the ability to handle occlusion better and faster. One drawback is the algorithms

**Fig. 8** A sample of bases and coefficients of SRC and LOFESS

might correspond to occluded regions and degrade the recognition performance.

explicitly. The constraint acts as a guidance for features training to concentrate on non-occluded facial parts. Figure 8 illustrates some bases and coefficient vectors of SRC (adopted from author's paper) and LOFESS (NMF-based) methods.

4 Experiments

4.1 Aleix-Roberts datasets

We evaluated the performance of LOFESS on the Aleix-Robert database [11] collected by Aleix Martine and Robert Benavente in Barcelona, 1999. There are 100 subjects, 50



Fig. 9 AR subset examples

men and 50 women, in the AR database. Each person has 2 images captured in 2 weeks apart for one facial status, there are 13 statuses in total.

This paper focuses on the disguised faces, so only AR-01 and AR-14 were chosen for training, AR-14, AR-08, AR-11, AR-21 and AR-24 for testing (Fig. 9). Each subset contains 100 images of 100 subjects captured in two week time apart in different conditions.

- AR-01, AR-14: neutral faces
- AR-08, AR-21: faces wearing sunglasses
- AR-11, AR-24: faces wearing scarves

Images are converted to 165×120 gray-scale in the pre-processing step.

4.2 Evaluation criteria

We performed extensive tests to evaluate the proposed method based on three criteria as summarized in Table 2.

4.2.1 Precision

This is the most popular criterion to evaluate the recognition rate, given by

$$P = \frac{\# \text{ correctly classified images}}{\# \text{ total classified images}}$$

AR-01 and AR-14 are used for training. AR-08, AR-11, AR-21 and AR-24 are for testing. Then results are compared with SLNMF [14] because both of them having the same experiment configuration.

4.2.2 Two week time recognition

Is LOFESS robust for recognizing a face two weeks later? In this test, only one image per subject in the subset AR-01 (Neu-1) was used for training and AR-08 (Sg-1), AR-11 (Sc-1), AR-13 (Neu-2), AR-21 (Sg-2), AR-24 (Sc-2) for testing. LOFESS was compared with SLNMF and RSC [5] based on the same testing configuration.

Table 2 Experiment summary

	Precision	2-week time	ROC
LOFESS	✓	✓	✓
SLNMF	✓	✓	
SRC		✓	

4.2.3 ROC curve

This curve reflects the correspondence between the true acceptance rate and false acceptance rate (plotted as the y and x axes, respectively) when recognition threshold is increased from 0 to 1. To our knowledge, there hasn't been any method addressing this problem has plotted the ROC curve for these AR datasets. We hope to provide another benchmark for later research.

4.3 Experimental results

4.3.1 Precision

Table 3 shows the recognition results on faces wearing sunglasses (a) and scarves (b). The main tables summarize recognition rates based on various local region sizes (in rows) and number of basis vectors n (in columns). Two sub-tables on the right and the bottom calculate the min, max and mean for each value of n and region size.

In detail, when the number of basis vectors varies from 10 to 300, the average precision increases from 68.8 to 91.17 % for the sunglasses subset and from 58.83 to 87.75 % for the scarf subset. But the rate is not stable if we look at the sub-table for region size, it goes up and down unpredictably. The wider the local region is, the smaller size of the basis is needed to achieve high recognition rate. This implies the optimal precision achieved with the appropriate choice of sufficient number of basis vectors and suitable regions size.

Tables 4 and 5 compared LOFESS and SLNMF under various number of bases. In case of recognizing targets with sunglasses, LOFESS outperforms SLNMF in all tests. But with scarf disguises, LOFESS is comparative with SLNMF in situations when only a few number of bases are allowed.

4.3.2 Two week time recognition

Optimal LOFESS recognition rate in each test is compared with SLNMF and RSC methods in this experiment as illustrated in Table 6. LOFESS and SLNMF used only one training image per subject in the subset AR-01, while RSC used up to 4 images in AR-01, AR-05, AR-06 and AR-07. Comparing with SLNMF, LOFESS outperformed in all tests. The main reason is LOFESS removes all occluded bases totally from

Table 3 Recognition precision on AR-08 and AR-21 (a), AR-11 and AR-24 (b)

n	Local Region Height / Image Height (%)						Min	Mean	Max
	3%	6%	9%	12%	15%	18%			
10	67	63	70.5	65	77	66	63	68.08	77
20	65.5	84	77.5	78.5	83	85.5	65.5	79	85.5
30	83	87	85.5	82.5	85	86.5	82.5	84.92	87
40	83.5	86	85	86.5	84	86	83.5	85.17	86.5
50	85.5	86.5	84.5	83.5	85.5	89.5	83.5	85.83	89.5
60	86.5	87.5	89.5	83.5	86.5	90.5	83.5	87.33	90.5
70	87.5	89	88.5	89.5	85.5	87.5	85.5	87.92	89.5
80	85.5	88.5	90.5	91	89.5	90	85.5	89.17	91
90	87.5	85.5	90	90	88	90	85.5	88.5	90
100	88.5	91.5	89.5	88.5	91.5	92.5	88.5	90.33	92.5
150	89.5	89	89.5	90.5	92.5	89.5	89	90.08	92.5
200	88	89.5	90.5	91.5	89.5	91	88	90	91.5
300	92	92	91	88.5	92	91.5	88.5	91.17	92
Min	65.5	63	70.5	65	77	66			
Mean	83.81	86.08	86.31	85.31	86.88	87.38			
Max	92	92	91	91.5	92.5	92.5			

(a)

n	Local Region Height / Image Height (%)						Min	Mean	Max
	3%	6%	9%	12%	15%	18%			
10	46.5	40.5	70	75	63	58	40.5	58.83	75
20	53.5	74.5	71.5	72	71	63	53.5	67.58	74.5
30	77	76	84	67.5	79.5	71	67.5	75.83	84
40	80	85.5	83.5	80.5	85.5	86.5	80	83.58	86.5
50	67	86.5	85	74	83	74	67	78.25	86.5
60	78.5	85	87	80	86	87.5	78.5	84	87.5
70	85.5	83	83.5	83	85.5	87	83	84.58	87
80	84.5	80.5	88.5	82	87	84.5	80.5	84.5	88.5
90	87	89	86.5	90	86	85.5	85.5	87.33	90
100	81	89	87.5	83	90	90	81	86.75	90
150	89	86.5	87.5	90.5	86.5	87.5	86.5	87.92	90.5
200	87.5	88	87	88	86	86	86	87.08	88
300	89.5	90	89.5	86.5	85	86	85	87.75	90
Min	46.5	40.5	70	67.5	63	58			
Mean	77.42	81.08	83.92	80.92	82.62	80.5			
Max	89.5	90	89.5	90.5	90	90			

(b)

Table 4 Recognition rate on the sunglasses dataset with various numbers of basis

Methods	Number of basic vectors			
	50	100	200	300
LOFESS	89.5	92.5	91.5	92
IS-LNMF	84	88	90	90

Table 5 Recognition rate on the scarf dataset with various numbers of basis

Methods	Number of basic vectors			
	50	100	200	300
LOFESS	86.5	90	88	90
S-LNMF	86	90	92	92

recognition. Meanwhile, SLNMF still tries to exploit bases partially corresponding to occluded area. In testing against

Table 6 Two week time period recognition rate (%) between LOFESS and SLNMF

Methods	Neu-2	Sg-1	Sg-2	Sc-1	Sc-2	# gallery images
LOFESS	80	91	67	91	61	1 image/person
S-LNMF	77	84	49	87	55	1 image/person

Table 7 Two week time period recognition rate (%) between LOFESS and RSC

Methods	Sg-1	Sg-2	Sc-1	Sc-2	# gallery images
LOFESS	91	67	91	61	1 image/person
RSC	94.7	91	80.3	72.7	4 images/person

RSC, the subset AR-11 (Sg-1) noticeably showed LOFESS reached a higher rate (91 %) even with one training image while RSC needed 4 images (80.3 %). The significant difference in performance between sunglasses and scarf datasets could be attributed to imprecise localization errors [24,25] (Table 7).

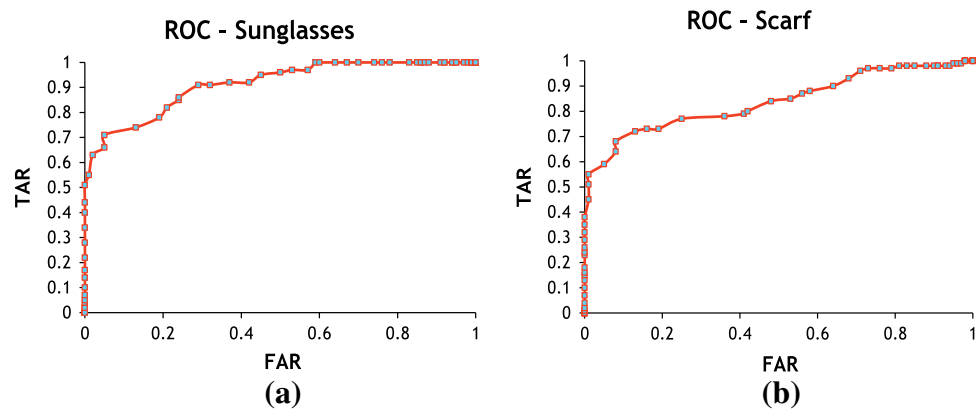
4.3.3 ROC curves

In Fig. 10, two ROC curves, which shows ratios between TAR and FAR, were plotted. Parameter values of $n = 300$ and region size = 3 % of image height were choice because this configuration gave the optimal performance among the experiments. In both subsets, the curves were above average line (the diagonal). However, when TAR = 1, the FAR was also quite high about 0.55 and 0.45, respectively. This is reasoned fusing just two images for training.

4.4 Parameter configuration and effects

4.4.1 Local region size r_1 , r_2 and the number of bases

In NMF-based methods, the number of bases could be infinite. LOFESS offers additional region size parameter. This allows more flexibility by tuning up both parameters for optimal solution. Here arises a question of how to find the best pairs of these values. Shatri and Levine [22] had a detailed survey on various number of basis vectors n from 10 to 200 with arbitrary region sizes. We performed experiments with the same bases number and varied the range of $[r_1, r_2]$ to occupy 3, 6, 9, 12, 15 and 18 percent of the image height. As presented in Table 3, region size tends to decrease while number of bases increases for the model to reach an saturated point. This implied the optimal recognition performance is reached when sufficient information is provided. A shortage or redundancy could downgrade the system.

Fig. 10 ROC curves for sunglasses (a) and scarf (b)**Table 8** Training time (min) for the sunglasses dataset

% image height	Number of bases								
	10	30	50	70	90	100	150	200	300
3	0.39	1.73	2.53	3.53	4.69	5.27	8.56	12.1	21.3
6	0.68	1.63	2.59	3.67	4.82	5.38	8.75	12.7	21.3
9	0.64	1.55	2.5	3.6	4.77	5.75	8.92	2.5	21.4
12	0.71	1.72	2.52	3.56	4.73	5.28	8.57	2.1	21.4
15	0.69	1.67	2.68	3.75	4.89	5.36	8.86	2.7	21.4
18	0.68	1.63	2.52	3.61	5.02	5.58	8.97	12.5	20.3

Table 9 Training time (min) for the scarf dataset

% image height	Number of bases								
	10	30	50	70	90	100	150	200	300
3	0.46	1.06	1.46	2.05	2.81	4.38	4.97	8.87	15.22
6	0.39	1.02	1.33	2.36	2.78	2.93	4.96	6.92	12.34
9	0.38	0.95	1.35	2.12	2.75	3	4.86	7.3	12.78
12	0.4	1	1.39	2.12	2.84	3.18	6.72	7.4	13.19
15	0.41	0.97	1.51	2.13	2.73	3.11	5.02	6.9	11.84
18	0.62	1.45	2.37	3.15	4.23	5.21	5.6	12.03	14.17

4.4.2 Training and matching time

In term of computation time, LOFESS converged after 200–500 iterations during training, which means the error function was almost stable. Detailed tables about training time (in min) corresponding to region size (in rows) and number of bases (in columns) are given in Tables 8 and 9. In return, the time for projecting a test image onto the LOFESS space and matching based on Euclidean distance is less than one second, which is ideal for real-time applications.

4.4.3 Occlusion form

Various occlusion forms should be handled differently due to their nature. For instance, a region occluded by scarf is wider than that by sunglasses. This loss in information some

how accounts for different results between types of occlusion. Basically, this is the common fact encountered in almost appearance-based approaches [14]. LOFESS allows a user to input a parameter telling which region should be discarded prior to the training phase. Then, the method automatically learns bases from non-occluded parts. In testing phase, all images are projected on these bases so occlusion is removed naturally.

5 Conclusions and future works

This paper presented the method Locality constrained feature representation for the disguised face recognition based on a small training set (LOFESS), which contains only one or two images per subject. By introducing spatially localized facial structure constraints, LOFESS effectively and efficiently captures prominent part-based features from non-occluded parts. Experiments showed this method is competitive with state of the art methods on AR datasets and can be extended to deal with other types of disguise, not just sunglasses or scarf. LOFESS is especially suitable for human supervising applications in which a suspect has his or her photos captured once or twice, such as identification (ID) or passport photos. Due to the constraint, features trained by NMF algorithm become more spatially localized and converge faster into expected facial regions. As a result, it obtains high recognition results even with very few training images.

Instead of prior knowledge from a user, LOFESS can be integrated with automatic occlusion detection algorithms. This is considered as our future work. After detecting occluded part, it is easily to exclude these regions and then follows the same process as presented in this paper. Alignment algorithms could be considered to enhance LOFESS robustness against time elapse. Moreover, how relationship between the optimal number of basis and the size of the extracted regions (the value r_1 and r_2) affects recognition performance also needs to be studied further.

Acknowledgments This research is funded by Vietnam National University HoChiMinh City (VNU-HCMC) under the project “Feature descriptor under variation condition for real-time face recognition application”, 2014.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Sinha, P.: Face recognition by humans: nineteen results all computer vision researchers should know about. *Proc. IEEE* **94**(11), 1948–1962 (2006)
2. Zhao, W., Chellapa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. *J. ACM Comput. Surv.* **35**(4), 399–458 (2003)
3. Azeem, A., Sharif, M., Raza, M., Murtaza, M.: A survey: face recognition techniques under partial occlusion. *Int Arab J Inform Technol* **11**(1), 1–10 (2011)
4. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2008)
5. Yang, M., Zhang, D., Yang, J., Zhang, D.: Robust sparse coding for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 625–632 (2011)
6. Chiang, C.C., Chen, Z.W.: Recognizing partially occluded faces by recovering normalized facial appearance. *Int. J. Innovative Comput. Inform. Control* **7**(11), 6210–6234 (2011)
7. Nguyen, M., Le, Q., Pham, V., Tran, T., Le, B.: Multi-scale sparse representation for Robust Face Recognition. *IEEE Third International Conference on Knowledge and Systems Engineering, KSE 2011*, Hanoi, Vietnam, October 14–17, pp. 195–199 (2011). ISBN 978-1-4577-1848-9
8. Rui, M., Hadid, A., Dugelay, J.: Improving the recognition of faces occluded by facial accessories. *IEEE International Conference on Automatic Face and Gesture Recognition and Workshops*, pp. 442–447 (2011)
9. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: *NIPS*, pp. 556–562 (2000)
10. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999)
11. Martine, A., Benavente, R.: The AR face database. <http://www2.ece.ohio-state.edu/aleix/ARdatabase.html> (2011)
12. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognitive Neurosci.* **3**(1), 71–86 (1991)
13. Matthews, I., Baker, S.: Active appearance models revisited. *Int. J. Comput. Vis.* **60**(2), 135–164 (2004)
14. Hyun, J.O., Lee, K.M., Lee, S.U.: Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. *Image Vis. Comput.* **26**(11), 1515–1523 (2008)
15. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. *European Conference on Computer Vision*, pp. 448–461 (2010)
16. Shen, L., Bai, L.: A review on gabor wavelets for face recognition. *Pattern Anal. Appl.* **9**, 273–292 (2006)
17. Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y.: Face recognition with contiguous occlusion using markov random fields. *International Conference on Computer Vision*, pp. 1050–1057 (2009)
18. Liao, S., Jain, A.K.: Partial face recognition: an alignment free approach. *International Joint Conference on Biometrics Compendium Biometrics*, pp. 1–8 (2011)
19. Lin, C.J.: On the convergence of multiplicative update algorithms for nonnegative matrix factorization. *IEEE Trans. Neural Netw.* **18**(6), 1589–1596 (2007)
20. Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. *Machine Learning*, pp. 1457–1469 (2004)
21. Hoyer, P.O.: Non-negative sparse coding. *Neural Networks for Signal Processing*, pp. 557–565 (2002)
22. Shastri, B.J., Levine, M.D.: Face recognition using localized features based on non-negative sparse coding. *Mach. Vis. Appl.* **18**(2), 107–122 (2007)
23. Li, S.Z., Hou, X.W., Zhang, H.J., Cheng, Q.S.: Learning spatially localized part-based representation. *IEEE Conference on Computer Vision Pattern Recognition*, pp. 207–212 (2001)
24. Ekenel, H.K., Stiefelhausen, R.: Why is facial occlusion a challenging problem. In: *International Conference on Biometrics* (2009)
25. Martinez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(6), 748–763 (2002)
26. Rentzeperis E., Stergiou A., Pnevmatikakis A., Polymenakos L.: Impact of face registration errors on recognition. In: *Artificial Intelligence Applications and Innovations*, pp. 187–194 (2006)
27. Chen, Y., Bao, H., He, X.: Non-negative local coordinate factorization for image representation. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 569–574 (2011)